

AD-A053 597

NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS F/G 12/1  
ON THE DISTRIBUTION OF QUANTILES OF RESIDUALS IN A LINEAR MODEL--ETC(U)  
MAR 78 R J CARROLL AFOSR-75-2796

UNCLASSIFIED

MIMEO SER-1161

AFOSR-TR-78-0819

NL

1 OF 1  
AD  
A053597



END  
DATE  
FILMED  
6-78  
DDC

AD A053597

AD No. ~~1~~  
DDC FILE COPY

ON THE DISTRIBUTION OF QUANTILES OF  
RESIDUALS IN A LINEAR MODEL

Raymond J. Carroll

Institute of Statistics Mimeo Series No. 1161

March, 1978

**DISTRIBUTION STATEMENT A**

Approved for public release;  
Distribution Unlimited

~~DEPARTMENT OF STATISTICS~~  
~~Chapel Hill, North Carolina~~

DDC  
RECEIVED  
MAY 5 1978  
B

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)  
NOTICE OF TRANSMITTAL TO DDC  
This technical report has been reviewed and is  
approved for public release IAW AFR 190-12 (7b).  
Distribution is unlimited.  
A. D. BLOSE  
Technical Information Officer

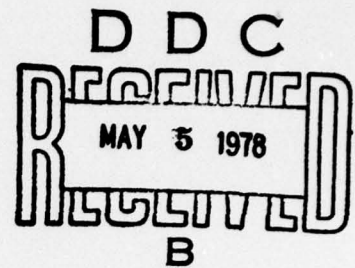


ON THE DISTRIBUTION OF QUANTILES OF  
RESIDUALS IN A LINEAR MODEL

by

Raymond J. Carroll\*

University of North Carolina at Chapel Hill



ABSTRACT

The representation for quantiles discussed by Bahadur (1966) and Ghosh (1971) is extended to quantiles from the residuals in a linear model. As an application, a robust test for symmetry is proposed.

AMS 1970 Subject Classifications: Primary 62G30; Secondary 62J05.

Key Words and Phrases: Quantiles, Order Statistics, Tests for Symmetry.

\* This work was supported by the Air Force Office of Scientific Research under contract AFOSR-75-2796.

DISTRIBUTION STATEMENT A

Approved for public release;  
Distribution Unlimited

## I. Introduction

Let  $y_1, \dots, y_n$  be independent and identically distributed with common distribution  $F$ , and let  $F(\xi) = p$ . Let  $k_n$  be a sequence of integers  $k_n = np + o(n^{1/2} \log n)$ , and let  $V_n$  be the  $k_n^{\text{th}}$  sample quantile. Bahadur (1966) and Ghosh (1971) have shown that if  $F_n$  is the empirical distribution function of  $y_1, \dots, y_n$ , then

$$(1) \quad n^{1/2} (V_n - \xi - (k_n/n - F_n(\xi))/F'(\xi)) \xrightarrow{P} 0.$$

Actually, Bahadur's result is an almost sure result, but (1) suffices for many statistical applications. Our goal is to extend (1) to the quantiles of the residuals in a simple linear model, using this to conduct a quick test for symmetry.

We consider the simple linear model  $z_i = x_i' \underline{\beta} + y_i$ , where  $y_1, \dots, y_n$  are i.i.d. with distribution  $F$  continuous and strictly increasing in a neighborhood of  $\xi$ , with  $F(\xi) = p$ . If  $\underline{\beta}_n$  is an estimate of  $\underline{\beta}$  (possibly the least squares estimate or one of the new robust estimates), the residuals are

$r_{in} = z_i - x_i' \underline{\beta}_n = y_i - x_i' (\underline{\beta}_n - \underline{\beta})$ . Let  $E_n(F_n)$  be the empirical distribution function of the residuals (errors), and choose  $a_n = n^{-1/2} \log n$ ,  $b_n = (\log n)^2$ . Let  $k_n$  be as above, let  $I_n = (\xi - a_n, \xi + a_n)$ , and define  $V_n$  to be the  $k_n^{\text{th}}$  order statistic of the residuals. We make the following assumptions:

(A1)  $F$  has two bounded derivatives in a neighborhood of  $\xi$ , and  $F'(\xi) = f(\xi) > 0$ .

(A2)  $a_n^{-1} (\underline{\beta}_n - \underline{\beta}) \rightarrow 0$  (a.s.)

SESSION for	
White Section	<input checked="" type="checkbox"/>
Buff Section	<input type="checkbox"/>
UNANNOUNCED	
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	Avail. and/or SPECIAL
A	

(A3) For some  $\delta > 0$ ,  $a_n^{1/2 - \delta} \max\{|x_i| : i=1, \dots, n\} \rightarrow 0$ .

(A4) There exists  $\underline{x}_0$  finite with  $(\log n)n^{-1} \sum_1^n (x_i - \underline{x}_0) \rightarrow 0$ .

**Theorem** Under conditions (A1) - (A4),

$$(2) \quad n^{1/2} |(V_n - \xi) - (k_n/n - F_n(\xi))/f(\xi) + \underline{x}_0' (\beta_n - \beta)| \rightarrow 0 \text{ (a.s.)}$$

**Remarks** Condition (A1) is the same as Bahadur's, while (A2) is quite reasonable. Condition (A3) guarantees that the design elements will not vary too wildly. The usual condition on the design matrix  $X$  is that  $n^{-1}X'X \rightarrow \Sigma$  (Drygas (1976)), and this implies that if there is an intercept term there is an  $\underline{x}_0$  with  $n^{-1} \sum_1^n (x_i - \underline{x}_0) \rightarrow 0$ , so that (A4) is only a slight strengthening of a standard condition, which in fact is reduced to  $n^{-1} \sum_1^n (x_i - \underline{x}_0) \rightarrow 0$  at the end of Section 2. Note that (A3) and (A4) are guaranteed in the problem considered by Bahadur and Ghosh.

## 2. Proof of the Theorem

We first consider simple linear regression without an intercept  $z_i = x_i\beta + y_i$ . The problems of simple and multiple linear regression with an intercept are easy consequences of the proof presented here.

Define the following processes:

$$G_n(x) = n^{-1} \sum_{i=1}^n \{I(r_i \leq x) - I(y_i \leq \xi) - (F(x + x_i(\beta_n - \beta)) - F(x))\}$$

$$H_n = n^{1/2} \sup\{|G_n(x)| : x \in I_n\}$$

$$W_n(s, t) = n^{-1/2} \sum_{i=1}^n \left\{ \begin{array}{l} I(y_i \leq \xi + a_n s + a_n t x_i) - I(y_i \leq \xi) \\ -F(\xi + a_n s + a_n t x_i) + F(\xi) \end{array} \right\}$$



where  $I(A)$  is the indicator of the event  $A$ . We note that from

(A1) - (A4),

$$(3) \quad n^{-1} \sum_{i=1}^n \{F(\xi + s a_n) - F(\xi + s a_n + t a_n x_i) + t a_n x_i f(\xi)\} = O(n^{-1/2})$$

uniformly for  $0 \leq |s|, |t| \leq 1$ .

Lemma 1 Under (A1) - (A4),  $H_n \rightarrow 0$  (a.s.) .

Proof of Lemma 1 Since  $r_{in} = y_i - x_i(\beta_n - \beta)$ , by (A2) it suffices to show that  $\sup \{|W_n(s, t)| : 0 \leq s, t, \leq 1\} \rightarrow 0$  (a.s.) .

For this, we may consider only  $x_i \geq 0$  by breaking the sum defining  $W_n$  into two parts, one with  $x_i \geq 0$  and the other over the terms  $x_i < 0$ . By (A1) as in Bahadur's proof, it also suffices to consider

points  $s, t = \eta_{r,n} = r/b_n$  ( $r = 0, \dots, b_n$ ). To see this, note that if  $|s - \eta_{r,n}| \leq b_n$ ,  $|t - \eta_{p,n}| \leq b_n$ , then

$$\begin{aligned} & |n^{-1} \sum_{i=1}^n \{F(\xi + s a_n + t a_n x_i) - F(\xi + \eta_{r,n} a_n + \eta_{p,n} a_n x_i)\}| \\ & = O(a_n b_n^{-1}) = O(n^{-1/2}). \end{aligned}$$

Now, by Bernstein's inequality (Hoeffding, eq. 2.13),

$$P_r\{|W_n(s, t)| > \epsilon\} \leq c_0 \exp\{-c n^{1/4}\}$$

for some constants  $c_0, c$  depending on  $\epsilon$ . Thus,

$$P_r\{|W_n(s,t)| > \epsilon \text{ for some } n \geq n_0, (s,t) = (\eta_{rn}, \eta_{pn})\} \\ \leq \sum_{n=n_0}^{\infty} c_0 b_n^2 \exp\{-c_1 n^{1/4}\} \rightarrow 0 \text{ as } n_0 \rightarrow \infty.$$

This completes the proof of the Lemma.  $\square$

Lemma 2 Almost surely as  $n \rightarrow \infty$ ,  $V_n \in I_n$ .

Proof of Lemma 2 We have

$$\Pr\{V_n \geq \xi + a_n\} = \Pr\left\{\sum_{i=1}^n I(y_i \leq \xi + a_n + x_i(\beta_n - \beta)) \leq k_n\right\}.$$

It suffices to show the existence of an  $\eta > 0$  for which

$Q_{n_0}(\eta) \rightarrow 0$ , where

$$Q_{n_0}(\eta) = \Pr\left\{\sum_{i=1}^n I(y_i \leq \xi + a_n + t a_n x_i) \leq k_n \text{ for some } 0 \leq t \leq \eta, n \geq n_0\right\} \\ = \Pr\{F_n(\xi + a_n) \leq k_n/n + n^{-1} \sum_{i=1}^n (F(\xi + a_n) - F(\xi + a_n + t a_n x_i)) \\ - n^{-1/2} W_n(1,t) \text{ for some } 0 \leq t \leq \eta, n \geq n_0\}.$$

By Lemma 1 and using equation (3),

$$Q_{n_0}(\eta) = \Pr\left\{F_n(\xi + a_n) - F(\xi + a_n) \leq -a_n f(\xi)(1 + t x_0) + o(a_n) \right. \\ \left. \text{for some } 0 \leq t \leq \eta, n \geq n_0\right\} + o(1).$$

Choosing  $\eta > 0$  small,  $Q_{n_0}(\eta) \rightarrow 0$  as in the proof of Bahadur's Lemma 2.  $\square$

Proof of the Theorem Since  $E_n(k_n) = k_n/n$ , Lemmas 1 and 2 may be applied to show that

$$n^{1/2} |F(V_n) - F(\xi) - (k_n/n - E_n(\xi))| \rightarrow 0 \text{ (a.s.)}.$$

Further,  $F(V_n) - F(\xi) = (V_n - \xi)f(\xi) + o(n^{-1/2})$  (a.s.) and

$$E_n(\xi) - F(\xi) = F_n(\xi) - F(\xi) + W_n(0, a_n^{-1}(\beta_n - \beta)) + n^{-1} \sum_{i=1}^n \{F(\xi + x_i(\beta_n - \beta)) - F(\xi)\}.$$



Since  $A_n^{-1}(\beta_n - \beta) \rightarrow 0$  (a.s.), the proof is complete by Lemma 1 and (A4).  $\square$

### 3. An Application

Hinkley (1975) suggested a method based on order statistics for estimating the parameter of a power transformation; his method is most applicable in the location parameter case. The Theorem can be used to construct a quick test for symmetry in more complicated models. Let  $p \neq 1/2$  and define  $V_n^{(p)}$  to be the  $[np]^{\text{th}}$  order residual. For  $F(\xi_p) = p$ ,

$$(4) \quad n^{1/2}(V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)} - (\xi_p + \xi_{1-p} - 2\xi_{1/2}))$$

has a limiting normal distribution (with mean zero and variance  $\sigma^2$ ) independent of the limit distribution of  $\beta_n$ . Under the hypothesis of symmetry,  $\xi_p + \xi_{1-p} = 2\xi_{1/2}$ , so one rejects the hypothesis if

$$(5) \quad n^{1/2}|V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)}|/\sigma$$

is too large. To estimate  $\sigma$  one needs estimates of the density at the quantiles  $\xi_p, \xi_{1-p}, \xi_{1/2}$ , which can be accomplished as in Bloch and Gastwirth (1968) by using the Theorem.

### References

1. Bahadur, R. R. (1966). A note on quantiles in large samples. Ann. Math. Statist. 37, 577-580.
  
- Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function. Ann. Math. Statist. 39, 1083-1085.
  
- Drygas, H. (1976). Weak and strong consistency of the least squares estimators in regression models. Z. Wahrscheinlichkeitstheorie Verw. Gebiete 34, 119-127.
  
- Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42, 1957-1961.
  
- Hinkley, D. V. (1975). On power transformations to symmetry. Biometrika 62, 101-111.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER <b>AFOSR-TR-78-0819</b>	2. GOVT ACCESSION NO. <b>14 MIMED</b>	3. RECIPIENT'S CATALOG NUMBER <b>SER-1161</b>	
4. TITLE (and Subtitle) <b>ON THE DISTRIBUTION OF QUANTILES OF RESIDUALS IN A LINEAR MODEL.</b>	5. TYPE OF REPORT & PERIOD COVERED <b>9 Interim rept.</b>		
7. AUTHOR(s) <b>Raymond J. Carroll</b>	6. PERFORMING ORG. REPORT NUMBER <b>Mimeo Series No 1161</b>		
9. PERFORMING ORGANIZATION NAME AND ADDRESS <b>University of North Carolina Department of Statistics Chapel Hill, NC 27514</b>	8. CONTRACT OR GRANT NUMBER(s) <b>15 / AFOSR-75-2796</b>		
11. CONTROLLING OFFICE NAME AND ADDRESS <b>Air Force Office of Scientific Rsch/NM Bolling AFB, DC 20332</b>	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>61102F 2304 A5</b>		
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	12. REPORT DATE <b>Mar 1978</b>		
	13. NUMBER OF PAGES <b>7</b>		
	15. SECURITY CLASS. (if this report) <b>UNCLASSIFIED</b>		
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE			
16. DISTRIBUTION STATEMENT (of this Report) <b>Approved for public release; distribution unlimited</b>			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  <b>quantiles, order statistics, tests for symmetry</b>			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  <b>The representation for quantiles discussed by Bahadur (1966) and Ghosh (1971) is extended to quantiles from the residuals in a linear model. As an application, a robust test for symmetry is proposed.</b>			